EDUCATION
RESEARCH SERVICE

ORIGINAL ARTICLE

Variables associated with successful performance on the National Board of Chiropractic Examiners Part IV examination

Igor Himelfarb, PhD, Bruce L. Shotts, DC, MS, John K. Hyland, DC, MPH, and Andrew R. Gow, DC, LAc

Objective: The objectives of this study were to (1) identify factors predictive of performance on the National Board of Chiropractic Examiners Part IV exam and (2) investigate correlations between the scores obtained in the Part I, Part II, Physiotherapy, and Part III exams and the Part IV examination.
Methods: A random sample of 1341 records was drawn from National Board of Chiropractic Examiners data to investigate the relationships between the scores obtained on the National Board of Chiropractic Examiners exams. A hierarchical multiple regression analysis related the performance on Part IV to examinee's gender, Part IV repeater status, and scores obtained on the Part I, Part II, Physiotherapy, and Part III exams.
Results: The analyses revealed statistical relations among all National Board of Chiropractic Examiners exams. The correlations between Part IV and Part I ranged from $r = .31$ to $r = .4$; between Part IV and Part II from $r = .34$ to $r = .45$. The correlation between Part IV and Physiotherapy was $r = .44$; between Part IV and Part III was $r = .46$. The strongest predictors of the Part IV score were found to be examinees' scores in Diagnostic Imaging, $\hat{\beta} = .19$, $p < .001$; Chiropractic Practice, $\hat{\beta} = .17$, $p < .001$; Physiotherapy, $\hat{\beta} = .15$, $p < .001$; and the Part III exam $\hat{\beta} = .19$, $p < .001$.
Conclusions: Performance on the National Board of Chiropractic Examiners Part IV examination is related to the performance in all other National Board of Chiropractic Examiners exams.

Key Indexing Terms: Chiropractic; Clinical Competency; Regression Analysis; Educational Measurement

J Chiropr Educ 2020;34(1):43–51 DOI 10.7899/JCE-18-28

INTRODUCTION

In early years, clinical competence was commonly assessed with essays and oral examinations.[1] Forty years ago, educators in the healthcare professions recognized a lack of student assessment of direct patient care. As a result, the objective structured clinical examination (OSCE) was proposed as an improvement to currently available assessment methods.[2] Since its introduction, the OSCE test of clinical competencies has become a critical part of training and certification of healthcare professionals.[3,4]

The OSCE is an assessment approach in which the components of clinical competence are assessed in a planned and structured way with attention being paid to objectivity of the examination.[5] OSCE candidates are observed and evaluated as they progress through a series of stations in which they interview or provide simulated treatment of standardized patients.[6] Various studies have been published on the use of OSCEs in health professions education, which reveal gaps that the present study may fill.

Studies From Other Healthcare Professions

Some studies have shown OSCE scores to be related to clinical curricula and "...general practice clinical attachment,"[7] which involves students working with clinicians and under their supervision in the last 2 years of medical training. The general attachment refers to training in the area of pediatrics and family medicine. The authors implemented a within subject pretest–posttest design using a sample of medical students who undertook an OSCE before and after their clinical education and attachment. The results of the study showed significant improvement in all OSCE station scores in the posttest phase.

Further, Peeraer et al[8] examined a sample of students who were given a questionnaire listing 182 basic clinical skills before and after their internship. The subjects were asked to report on the number of times they performed each skill during their internship. A 14-station OSCE tool was used to assess their basic clinical skills procedures as taught during the first 5 years of the medical curriculum. The results contradicted the Townsend et al[7] study—no significant difference was shown in the postinternship OSCE scores when compared to preinternship scores.

Yet another study examined a sample of medical students following a 15-week "attachment" in pediatrics and child health, general practice, and dermatology in participants' 2nd clinical year.[9] These students were assessed on a 10-station OSCE and scores were used to fit a confirmatory factor analytic model. The analysis showed a significant relation between the OSCE scores and the assessment of clinical skills across all specialties.

Studies From the Chiropractic Field

Recently, several studies examined the relationship between academic achievement in science courses and scores received on various parts of the National Board of Chiropractic Examiners (NBCE) exams. McCall and Harvey[10] studied the relationships between Part I and Part II scores, students' incoming grade point average (GPA), and their course-related GPAs. The goal of the researchers was to evaluate the factors predictive of successful performance on Parts I and II.

Two models were constructed—first to predict scores obtained on the Part I exam, and second to predict scores obtained on the Part II exam. Their results revealed close relationships between the GPA based on the performance in Part I-related courses (Part I GPA) and the scores obtained on the Part I exam. In fact, based on the first model constructed by McCall and Harvey, 60% of the variability in Part I scores is explained by the Part I GPA. Based on these results, the authors claim to have established criterion validity for the Part I scores for the population of Logan University students. In the second model, predicting the Part II scores, 68% out of a total 75% of explained variability was accounted by the linear combination of the Part I GPA and Part I scores. Thus, the claim of establishing criterion validity for Part II is less convincing. Further, while McCall and Harvey[10] derived informative conclusions, their study suffered from a major limitation—all data used in the study were obtained from Logan University, which raises the question of generalizability of the findings.

In a separate study,[11] researchers assessed the relations between entering GPA, academic performance at Sherman College of Chiropractic, the success in a test preparatory course, and scores obtained on NBCE's Part I exam. The researchers found strong correlations between the NBCE scores and entry GPA ($r = .46$); NBCE scores and academic performance (correlations ranged from $r = .51$ to $r = .76$). The results are not surprising as the NBCE periodically studies the chiropractic colleges' curricula using the Delphi method[12] as a part of its exam development efforts. The authors continued with a construction of a linear model regressing Part I scores on the averages obtained in anatomy and chemistry classes. Both variables emerged as successful predictors. The researchers excluded the 4 other domains of Part I from the regression equation due to the violation of linearity assumption limiting the predictive qualities of their model. Further limitations of their study were the very small ($n = 24$) sample and the fact that all participants were drawn from the student population at a single chiropractic college.

Knowledge acquired during a formal educational process at a chiropractic college as well as extracurricular participation had been found to be influential on the choice of chiropractic technique (a domain tested by the Part IV exam) used during internship and later in practice.[13] Researchers surveyed 164 students inquiring about the relation between the preferred chiropractic technique and their educational curriculum. They found that, "students appear to have the same practice technique preferences as practicing chiropractors. The chiropractic technique curriculum and the students' experience with chiropractic practitioners seem to have the greatest influence on their choice of chiropractic technique for future practice."[13] The researchers stated that students formed their chiropractic technique preferences based on the chiropractic curriculum, technique clubs, technique seminars, and chiropractic practitioners they encounter.

Aims

Considering the fact that Part IV is a gatekeeper for chiropractic licensure, it is essential to identify, in a more generalizable way, the factors that are predictive of successful performance on the test. The research emerging from the medical field has established a connection between OSCE performance and clinical training. However, less information is available on the connection between OSCE scores and performance on other types of assessment or a connection to the nonclinical curricula. Furthermore, to the best of our knowledge, this is the first study to examine the correlates of successful performance on the Part IV exam, and to relate the Part IV scores to performance on other NBCE exams. The primary goal of this study was to establish factors predictive of the performance on the Part IV exam. Additionally, we investigated relationships between the scores obtained in the Part I, Part II, Physiotherapy, Part III exams and the Part IV examination.

The research questions related to the objectives of this study were as follows:

1. What is the relationship between the performance on Part I, Part II, Physiotherapy, Part III, and Part IV exams?
2. What are the important predictors of performance on Part IV exam?
3. Are there gender differences in performance on the NBCE exams?
4. Are there differences in exam performance between examinees who pass Part IV on the first attempt and those examinees who repeat the test?

METHODS

The NBCE Part IV OSCE

Part IV is a performance-based exam, which consists of case history, physical exam, and orthopedic and neurologic stations. In addition, the Part IV exam incorporates chiropractic technique stations where examinees are evaluated on adjusting techniques. During the exam, the

candidates progress through the stations where they interview standardized patients and perform diagnostic procedures. Following patient encounters, the examinees are required to answer questions based on the information collected from the simulated patients.

Initially, the Part IV exam was built around 3 domains: Diagnostic Imaging, Chiropractic Technique, and Case Management. The test was scored using classical test theory methods until 2018. Starting with the May 2018 administration, Part IV has been scored using item response theory (IRT) models.[14,15] As a prerequisite to fitting IRT models, the number of domains underlying the Part IV exam was reexamined. While the Part IV exam continues to follow previous test plans, we found it to be a 4-dimensional exam. For scoring purposes, the Case Management portion was split into 2 domains: Patient Encounters and Post Encounter Probes.

The NBCE Written Exams

The NBCE Written Exams consist of the Part I, Part II, Physiotherapy, and Part III examinations. Part I is a test of basic sciences, and consists of 6 domains: General Anatomy, Spinal Anatomy, Physiology, Chemistry, Pathology, and Microbiology. Part II is a test of clinical sciences with the 6 following domains: General Diagnosis, Neuromusculoskeletal Diagnosis, Diagnostic Imaging, Principles of Chiropractic, Chiropractic Practice, and Associate Clinical Sciences. The Part III exam is a written clinical competency test consisting of 9 clinical areas: Case History, Physical Examination, Neuromuskuloskeletal Examination, Diagnostic Imaging, Clinical Laboratory and Special Studies, Diagnosis or Clinical Impression, Chiropractic Techniques, Supportive Interventions, and Case Management. Physiotherapy is an elective exam usually taken in conjunction with other NBCE exams, but not required by all states.

Ethics

The proposal for this study was reviewed and approved by the NBCE institutional review board (IRB). The study used secondary data for statistical analyses; however, the authors and the IRB committee wanted to assure the protection of all examinees' identities. The goal was to develop and verify protocols that would prevent relating the data used for the research or published results back to the examinees.

Data and Procedures

Data Collection

The data for this study were obtained from NBCE's internal data depository. We used the standard sample size formula for estimating sample percentage given by the following:

$$n = \frac{z^2(pq)}{se^2}$$

where n is the sample size, z is the standard score associated with a chosen level of confidence, p is the estimated percent in the population, $q = 100 - p$, and se is the accepted sample error.[16] Based on the calculations, 30% random sample ($n = 2083$) was drawn from the records of examinees containing scaled scores for the Part I, Part II, Physiotherapy, Part III, and Part IV examinations for the 2 years prior to the May 2018 Part IV administration using "Select Cases: Random Sample" function within SPSS 24.[17] This functionality generates a random sample of approximately the specified percentage of cases. The routine makes an independent pseudorandom decision for each case while selecting the approximating of the prespecified sample percentage.

Data Preparation

To prepare the data for the study, a preliminary analysis of descriptive statistics and frequencies was conducted. Not every test taker took all 5 tests, resulting in some missing values in the data. We learned that data were missing at random, therefore, the list-wise deletion method was determined to be appropriate. List-wise deletion removes all information for a case that has 1 or more missing values.[18] The final analysis sample contained $n = 1341$ examinees with no missing values on any measure.

The original file (file 1) included test takers' names, which were replaced by generic examinee IDs. Next, the names were removed from the analysis file (file 2). When file 2 was determined final, file 1, with examinees' names, was deleted making it impossible to relate file 2 to the original database.

Measures

Gender

Currently, the NBCE does not collect test takers' gender during the registration process. However, we felt that gender is an important variable to be included in this study, so we used examinee's first and/or middle names to "guess" gender. For test takers with foreign names we first determined the ethnic/cultural background for the test taker and then consulted with individuals who were familiar with that culture. We were able to "guess" gender for all 1341 test takers. Gender was coded 0 (if female) and 1 (if male).

Part IV Repeater Status

The NBCE maintains records for examinees who fail their initial Part IV exam and choose to repeat it. This variable was coded 0 (if nonrepeater) and 1 (if repeater).

Part IV Scores

The scores for Part IV are produced based on the examinees' item responses. The raw responses are scored according to predetermined rubrics; the responses are calibrated, equated, and scaled. The scale for the Part IV ranges from 125 to 800 with $M = 500$, SD = 100.

Part I and Part II Scores

The psychometric processes of score production for the Part I and Part II exams are identical. Currently, each domain is scored independently. The raw responses are scored dichotomously (1 if correct and 0 if incorrect) and

Table 1 - Descriptive Statistics for Test Scores

Domain	*n*	Mean	SD
Part I			
General Anatomy	1341	496.6	81.9
Spinal Anatomy	1341	497.0	85.4
Physiology	1341	490.4	79.2
Chemistry	1341	497.8	79.8
Pathology	1341	515.5	89.0
Microbiology	1341	507.6	82.3
Part II			
General Diagnosis	1341	508.0	84.0
Neuromusculoskeletal Diagnosis	1341	506.4	87.5
Diagnostic Imaging	1341	502.6	86.6
Principles of Chiropractic	1341	512.0	85.6
Chiropractic Practice	1341	514.4	93.8
Associated Clinical Sciences	1341	518.3	88.4
Physiotherapy	1341	500.5	81.7
Part III	1341	496.1	81.7
Part IV	1341	519.9	68.1

are calibrated using the IRT models. The scores are equated and then scaled. The scale for the Part I and Part II exams ranges from 125 to 800 with $M = 500$, SD = 100.

Physiotherapy and Part III

The scores for Physiotherapy and Part III are produced based on the examinees' item responses. The raw responses are scored, calibrated using IRT models, equated, and scaled. The scales for the Physiotherapy and Part III scores range from 125 to 800 with $M = 500$, SD = 100.

Analytic strategy

Correlations and Mean Differences

A correlation matrix was constructed to examine bivariate relations between gender, Part IV repeater status, and the scores of Part I, Part II, Physiotherapy, Part III, and Part IV. Next, a set of independent-sample t tests assessed the mean differences in Part I, Part II, Physiotherapy, and Part III scores between the examinees who repeated the Part IV exam and those who did not. Effect sizes were estimated for the t tests to evaluate the strength of statistical conclusions.

Prediction

The hierarchical multiple regression model,[19] the predictive analysis, was estimated to offer a comprehensive model predicting the Part IV scores. In the hierarchical multiple regression, predictors were included in the model in the hypothesized order specified by the researchers.

Hierarchical multiple regression is considered an appropriate analytic tool when variance of the dependent variable is being explained by sets of predictors.[20] The total variance in the dependent variable explained by the model is disaggregated according to the contribution of each variable block (step in the hierarchy of the model). The change in R^2 provides an exact contribution of each step, while controlling for the contribution of all previous steps.

Table 2 - Descriptive Statistics for Gender and Part IV Repeater Status

	n	Percent
Gender		
Female	529	39.4
Male	812	60.6
Repeater status		
Non-repeater	1154	86.1
Repeater	187	13.9

Regression Model

The formulaic representation of the estimated model:

$$\begin{aligned} PartIV_i = {} & \beta_0 + \beta_1 d(Male)_i + \beta_2 d(\text{Repeater})_i + \beta_3 (GEA)_i \\ & + \beta_4 (SPA)_i + \beta_5 (PHY)_i + \beta_6 (CHE)_i + \beta_7 (PAT)_i \\ & + \beta_8 (MIC)_i + \beta_9 (GED)_i + \beta_{10} (NMS)_i \\ & + \beta_{11} (DIM)_i + \beta_{12} (PRI)_i + \beta_{13} (PRA)_i \\ & + \beta_{14} (ACS)_i + \beta_{15} (PHT)_i + \beta_{16} (Part\ III)_i + e_i \end{aligned}$$

where $PartIV_i$ is the Part IV score for an examinee $i = 1, \ldots,$ n; β_0 is the intercept, and $\beta_1, \ldots, \beta_{16}$ are regression coefficients (slopes) associated with the predictors in the model. The term $d(\cdot)$ is the indicator for the parenthesized predictor; that is, $d = 1$ if the examinee has the characteristic in question and is 0 otherwise. The e_i is the error term representing the difference between the score predicted by the model and the observed score in the data.

We make the following assumptions for the model:

$$PartIV_i \sim N(\mu; \sigma^2)$$

To test this assumption, the distribution of Part IV scores was examined. The scores were symmetrically distributed around the mean; skewness was .04, SE = .07, and the value of kurtosis was −.56, SE = .14. Based on these estimates, we concluded that the assumption of normality was upheld. The following was assumed for the distribution of the error term:

$$e_i \overset{iid}{\sim} N(0, \sigma^2)$$

The errors were normally distributed, with the expectation of 0, and a constant variance that does not depend on predictors.

RESULTS

Descriptive Statistics and Percentages

The descriptive statistics for test scores are presented in Table 1. The scaled-score averages for all exams approximated the mean of the scale, $M = 500$. The descriptive statistics for the sample are presented in Table 2.

Bonferroni Correction

The Bonferroni adjustment[21] was made to control for inflated Type I error due to multiple analyses on the same dependent variable. The correction was made to the p values used for evaluation of statistical significance. We used the following formula to perform Bonferroni

correction:

$$\alpha_{critical} = 1 - (1 - \alpha_{altered})^k$$

where k is the number of comparisons. As the result, all decisions of statistical significance were made at the $\alpha = .001$ level. The confidence intervals associated with t tests and the regression are presented at the 99.9% level.

Correlations

Gender

First-order correlations are presented in Table 3. Gender (reference group is female) showed significant negative correlation with Pathology, $r = -.11$, $p < .001$; Microbiology, $r = -.11$, $p < .001$; General Diagnosis, $r = -.14$, $p < .001$; Diagnostic Imaging, $r = -.07$, $p < .001$; and Associated Clinical Sciences, $r = -.17$, $p < .001$. These correlation estimates indicate that, on average, female examinees perform better than male examinees in the areas listed above. Gender revealed positive correlation with Principles of Chiropractic, $r = .08$, $p < .001$, suggesting that men perform better than women in that domain.

Part IV Repeater Status

Part IV repeater status (reference group is nonrepeater) exhibited moderate negative correlations with test scores in Part I, Part II, Physiotherapy, Part III, and Part IV. The correlation estimates ranged from $r = -.16$, $p < .001$ to $r = -.26$, $p < .001$. These results show that, on average, Part IV repeaters tend to get lower test scores in all NBCE assessment programs.

Part I (Cronbach $\alpha = .92$)

All domains of Part I were found to be interrelated. The estimates of bivariate correlations were all high and positive, ranging from $r = .56$, $p < .001$ to $r = .78$, $p < .001$. The high positive correlations between the domains of the test provide evidence of convergent validity and internal consistency of the Part I domains. A measure is judged to have convergent validity when it is positively correlated with other types of measures of the same or similar constructs.[22] The internal consistency is the extent to which the components of an instrument are interrelated.[23]

Part II (Cronbach $\alpha = .89$)

All domains of Part II were found to be interrelated, showing high statistically significant bivariate relations. The correlation estimates ranged from $r = .47$, $p < .001$ to $r = .71$, $p < .001$. Similar to Part I, these correlations provide evidence of convergent validity and internal consistency.

The correlation estimates between the domains of Part I and domains of Part II were all positive and significant. The estimates ranged from $r = .43$, $p < .001$ to $r = .7$, $p < .001$.

Physiotherapy, Part III, and Part IV

Physiotherapy revealed a significant relationship with Part I, with correlation estimates ranging from $r = .39$, $p < .001$ to $r = .61$, $p < .001$; Part II, correlation estimates

Table 3 - First-Order Correlations for Gender, Part IV Repeater Status, and Test Scores

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Gender	–																
2. Repeater	.04	–															
3. General Anatomy	.00	−.19**	–														
4. Spinal Anatomy	−.03	−.21**	.78**	–													
5. Physiology	.04	−.19**	.76**	.77**	–												
6. Chemistry	−.05	−.19**	.67**	.7**	.73**	–											
7. Pathology	−.11**	−.20**	.63**	.64**	.68**	.66**	–										
8. Microbiology	−.11**	−.16**	.56**	.57**	.58**	.61**	.66**	–									
9. General Diagnosis	−.14**	−.22**	.6**	.6**	.62**	.62**	.68**	.61**	–								
10. Neuromusculoskeletal Diagnosis	−.03	−.24**	.63**	.66**	.61**	.6**	.64**	.5**	.64**	–							
11. Diagnostic Imaging	−.07	−.24**	.63**	.64**	.61**	.58**	.66**	.55**	.66**	.68**	–						
12. Principles of Chiropractic	.08**	−.22**	.65**	.66**	.7**	.58**	.6**	.48**	.6**	.62**	.68**	–					
13. Chiropractic Practice	.04	−.24**	.57**	.57**	.59**	.49**	.52**	.42**	.52**	.58**	.59**	.66**	–				
14. Associated Clinical Sciences	−.17**	−.19**	.52**	.51**	.54**	.52**	.61**	.58**	.71**	.53**	.58**	.51**	.47**	–			
15. Physiotherapy	.04	−.25**	.58**	.54**	.61**	.49**	.51**	.39**	.51**	.55**	.57**	.65**	.61**	.46**	–		
16. Part III	−.02	−.26**	.59**	.59**	.57**	.54**	.58**	.48**	.61**	.67**	.65**	.63**	.58**	.56**	.53**	–	
17. Part IV	−.06	−.19**	.40**	.4**	.38**	.37**	.39**	.31**	.37**	.43**	.45**	.39**	.44**	.34**	.44**	.46**	–

** $p < .001$.

Table 4 - Mean Differences in Test Scores as a function of Part IV Repeater Status

Variable	Non-Repeater		Repeater				99.9% CI for Δ(*M1-M2*)	
	Mean	SD	Mean	SD	*t* test	Cohen *d*	Lower	Upper
General Anatomy	502.9	82.9	457.8	62.9	7.1**	0.6	24.1	66.0
Spinal Anatomy	504.2	86.7	453	60.3	7.8**	0.7	29.4	72.9
Physiology	496.6	80.5	452.7	58.6	7.1**	0.6	23.6	64.1
Chemistry	504	80.6	460.1	62.6	7.2**	0.6	23.5	64.3
Pathology	522.7	89.9	471	68.6	7.5**	0.7	29.1	74.4
Microbiology	512.8	82.7	475.9	72.3	5.7**	0.5	15.7	58.0
General Diagnosis	515.7	84.3	462.3	66.0	8.1**	0.7	31.5	75.2
Neuromusculoskeletal Diagnosis	514.9	88.5	455.5	60.0	8.5**	0.8	36.3	82.5
Diagnostic Imaging	511.4	87.6	451.6	59.1	8.7**	0.8	37.1	82.4
Principles of Chiropractic	519.7	86.7	466.3	62.2	7.8**	0.7	30.8	76.0
Chiropractic Practice	523.4	94.7	460	66.0	8.4**	0.8	38.6	88.2
Associated Clinical Sciences	525.1	88.6	478.01	75.3	6.7**	0.6	23.8	70.3
Physiotherapy	509.5	81.4	453.9	66.2	8.1**	0.7	33.0	78.1
Part III	505	82.4	447.3	56.6	8.2**	0.8	34.4	81.0

** $p < .001$; two-tailed.

ranging from $r = .46$, $p < .001$ to $r = .65$, $p < .001$; Part III, $r = .53$, $p < .001$; and Part IV, $r = .44$, $p < .001$.

Statistically significant correlations were estimated between Part III and the domains of Part I, correlations ranging from $r = .48$, $p < .001$ to $r = .59$, $p < .001$; the domains of Part II, correlations ranging from $r = .56$, $p < .001$ to $r = .67$, $p < .001$; and Part IV, $r = .46$, $p < .001$.

The correlation estimates between Part IV and the domains of Part I, and between Part IV and the domains of Part II were all positive and statistically significant. The estimates ranged from $r = .31$, $p < .001$ to $r = .4$, $p < .001$ for Part I, and from $r = .34$, $p < .001$ to $r = .45$, $p < .001$ for Part II.

Mean Differences

Independent samples *t* tests evaluated the mean differences in Part I, Part II, Physiotherapy, and Part III scores as a function of being a Part IV repeater. A systematically better performance in all of the assessments was found for examinees who did not need to repeat the Part IV exam. Large effect sizes were documented for all mean differences assessed.[24] The results for the tests of mean differences are presented in Table 4.

Predictive Model

A 4-step hierarchical linear multiple regression model was estimated to test the contribution of each predictor as well as each block of predictors on Part IV scores, a technique that is useful for evaluation of contribution of predictors above and beyond those previously entered in the model.[21] The gender and Part IV repeater status were controlled for in step 1. In step 2, the 6 domains of Part I were included in the model. Step 3 incorporated the 6 domains of Part II, and in step 4 Physiotherapy and Part III were included. Despite the number of predictors in the model, the problem of multicollinearity was not encountered, as all predictors displayed acceptable levels of tolerance, ranging from .51 to .85. The results of the regression model are presented in Table 5.

The gender and Part IV repeater status revealed significant prediction as a block explaining 6% of variability in Part IV scores, $F_{change(2,1339)} = 22.85$, $p < .001$. The 6 domains of Part I revealed predictive qualities as a block explaining an additional 13% of variability in Part IV scores, $F_{change(8,1333)} = 21.19$, $p < .001$. The block of Part II domains emerged as a significant predictor, explaining an additional 9% of the variability in Part IV scores, $F_{change(14,1327)} = 20.01$, $p < .001$. Finally, Physiotherapy and Part III were predictive as a block, explaining 3% of the variability in Part IV scores after accounting for gender, Part IV repeater status, Part I, and Part II, $F_{change(16,1325)} = 20.04$, $p < .001$.

The final model, incorporating all predictors, explained 31% of the total variance in Part IV scores. After Bonferroni correction, gender and repeater status did not reveal statistical significance. Although Part I scores were found to be significant as a block, none of the individual domains revealed statistical significance in predicting the Part IV scores. In block 3 (Part II), the Diagnostic Imaging, $\hat{\beta} = .19$, $p < .001$, and Chiropractic Practice, $\hat{\beta} = .17$, $p < .001$, scores were identified as predictors of Part IV scores. Finally, both Physiotherapy, $\hat{\beta} = .15$, $p < .001$, and Part III, $\hat{\beta} = .19$, $p < .001$ emerged as significant predictors.

DISCUSSION

Previous research has shown a connection of OSCE scores to clinical studies,[25,26] other standardized tests,[27] and various nonclinical learning outcomes in several specialties and disciplines.[28] However, virtually no information was available on the Part IV, the OSCE exam of the chiropractic field. This current study was designed to comprehensively scrutinize the predictors of Part IV exam scores and the relationships between Part IV scores and the scores on other NBCE's exams.

This research has provided new information about the NBCE's exams. Particularly, we showed that, on average, female test takers tend to score better than male examinees

Table 5 - Hierarchical Multiple Regression Predicting Part IV Scores

Predictor	At Step			Final Model			99.9% CI for *B*	
	Δ R^2	R^2	Multiple *R*	*B*	SE	β	Lower	Upper
Step 1: Gender and repeater status	.06**	.06**	.24**					
Gender				−9.04	4.43	−.07	−29.64	10.59
Repeater				−9.52	4.09	−.05	−23.70	5.62
Step 2: Part I scores	.13**	.19**	.44**					
General Anatomy				.02	.04	.03	−.12	.17
Spinal Anatomy				.02	.04	.03	−.13	.17
Physiology				−.06	.05	−.08	−.23	.10
Chemistry				.05	.04	.06	−.80	.19
Pathology				.01	.04	.02	−.11	.14
Microbiology				−.04	.04	−.05	−.16	.08
Step 3: Part II scores	.09**	.28**	.53**					
General Diagnosis				−.04	.04	−.05	−.18	.09
Neuromusculoskeletal Diagnosis				.06	.04	.08	−.70	.19
Diagnostic Imaging				.14	.04	.19**	.02	.27
Principles of Chiropractic				−.06	.04	−.07	−.20	.08
Chiropractic Practice				.12	.03	.17**	.02	.22
Associated Clinical Sciences				−.03	.03	−.04	−.14	.09
Step 4: Physiotherapy and Part III scores	.03**	.31**	.56**					
Physiotherapy				.12	.04	.15**	.05	.24
Part III				.16	.04	.19**	.03	.29

** $p < .001$.

Reference group for gender is female and for Part IV repeater status is nonrepeater.

in Pathology, Microbiology, General Diagnosis, Associated Clinical Sciences, and the Part IV. Yet, the average of scores for males was higher in Principles of Chiropractic. Prior to the analyses, we hypothesized that the domains of all other exams would emerge as correlates of the Part IV scores. The current study supports the hypothesized relationships between all domains of the Part I and Part II, Physiotherapy, Part III, and Part IV by revealing strong positive correlations. Additionally, the Part IV repeater status negatively correlated with scores in all other exams.

A specific forte of this study is the origin and the breadth of the data. While authors of several other studies[10,11,13] attempted to relate the NBCE scores to aspects of their research interests, the research was limited by the availability of data, which was often confined to a single chiropractic college. On the other hand, this study used global data across all chiropractic colleges that participate in the NBCE's testing programs. Therefore, the results of this study are more externally valid for the general population of examinees. Moreover, in this study we randomly sampled from the population of test takers, arriving at a substantial sample size. The random sampling reduced the degree of possible inaccuracy, which may be caused by biases associated with self-selected samples.

Limitations

The results of this study should be interpreted in view of several limitations. For the purposes of this research, gender was "assigned" to the participants based on their names (first name and middle name) and by looking at the examinee's picture included with the NBCE's initial application. Therefore, gender-related conclusions should be interpreted with caution. The NBCE now collects demographic information from the candidates, which may allow future validation of this study's results using self-reported gender.

Future research should determine the relationships to other imperative factors predictive of Part IV performance but omitted by the current study. Our study's design used a nonexperimental approach evaluating cross-sectional variables. As such, the study lacks random assignment and a formal control group; thus, causal relationships may not be established between the predictors and the outcomes.[29]

In this research, we used the data available to the NBCE to construct and test a model predicting Part IV performance. The practical information may benefit both chiropractic students and educators by providing factors to focus on and early warning signs to increase the likelihood of achieving a passing Part IV score. We hope that the chiropractic community finds this study informative and contributing to test takers' success.

CONCLUSION

We sought to study the relations among the scores obtained on prelicensure NBCE exams. In the latter portion of the study, we wanted to identify factors predictive of the Part IV score. The results revealed interrelation among domain scores of the Part I and Part II. Statistical association was revealed at the test level—the Part I, as a test, correlated with Part II and Part III. The scores obtained on the Part IV exam significantly correlated with every domain of the Part I, Part II, Physiotherapy, and Part III.

The hierarchical multiple regression allowed for examination of the unique effects of each of the exams while controlling for available demographic characteristics in the sample. While controlling for multicollinearity, all tests included in the analysis were found to be explanatory of the Part IV scores' variability. The 2 strongest predictors of the Part IV performance were the scores obtained on the domain of Diagnostic Imaging (Part II) and the scores in Part III. These findings are both logical and in line with the previous research. Additionally, the scores attained in Physiotherapy emerged as a strong predictor of scores in Part IV.

FUNDING AND CONFLICTS OF INTEREST

This work was funded internally. The authors are employees of the NBCE.

About the Authors

Igor Himelfarb is the director of Psychometrics and Research at the National Board of Chiropractic Examiners (901 54th Avenue, Greeley, CO 80634; ihimelfarb@nbce.org). Bruce Shotts is the director of Written Examinations at the National Board of Chiropractic Examiners (901 54th Avenue, Greeley, CO 80634; bsotts@nbce.org). John Hyland is a consultant in the Department of Psychometrics and Research at the National Board of Chiropractic Examiners (901 54th Avenue, Greeley, CO 80634; jhyland@nbce.org). Andrew Gow is the director of Practical Testing, Research and Development at the National Board of Chiropractic Examiners (901 54th Avenue, Greeley, CO 80634; agow@nbce.org). Address correspondence to Igor Himelfarb, 901 54th Avenue, Greeley, CO 80634; ihimelfarb@nbce.org. This article was received September 16, 2018, revised January 29, 2019, and March 11, 2019, and accepted March 19, 2019.

Author Contributions

Concept development: IH, JH, BS. Design: IH. Supervision: BS, JH, AG. Data collection/processing: IH. Analysis/interpretation: IH, BS, JH, AG. Literature search: IH, BS, AG. Writing: IH, JH. Critical review: BS, JH, AG.

REFERENCES

1. Newble D. Techniques for measuring clinical competence: objective structured clinical examinations. *Med Educ*. 2004;38:199–203.
2. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. *BMJ*. 1975;(1):447–451.
3. Baransky B, Etzel SI. Medical schools in the United States, 2009-2010. *JAMA*. 2010;304:1247–1254.
4. Hillard RI, Taller SE. The use of an objective structured clinical examination with postgraduate residents in pediatrics. *Arch Pediatr Adolesc Med*. 1998;152:74–78.
5. Harden RM. What is an OSCE? *Med Teach*. 1988; 10(1):19–22.
6. Harden RM, Gleeson F. Assessment of clinical competence using an objective structured clinical examination (OSCE). *BMJ*. 1979;13(1):41–54.
7. Townsend AH, Mcllvenny S, Miller CJ, Dunn EV. The use of an objective structured clinical examination (OSCE) for formative and summative assessment in a general practice clinical attachment and its relationship to final medical school examination performance. *Med Educ*. 2016;35(84):841–846.
8. Peeraer G, Muijtjens AMM, De Winter DY, Remmer R, Hendrickx K. Unintentional failure to assess for experience in senior undergraduate OSCE scoring. *Med Educ*. 2008;42:669–675.
9. Reece A, Chung EMK, Gardiner RM, Williams SE. Competency domains in an undergraduate Objective Structured Clinical Examination: their impact on compensatory standard setting. *Med Educ*. 2008;42: 600–606.
10. McCall AR, Harvey RD. Predictors of performance on the National Board of Chiropractic Examiners Parts I and II. *J Chiropr Educ*. 2014;28(1):9–15.
11. Kenya AWM, Kenya HM, Hart J. Correlation between academic performance and NBCE part I scores at a chiropractic college. *J Chiropr Educ*. 2013; 27(1):27–32.
12. Gordon TJ, Helmer O. Report on a Long-Range Forecasting Study. Santa Monica, California: RAND Corporation; 1964 Sep. Report No.: AD0607777.
13. Sikorski DM, Veettil AK, Tobias GS. The influence of curricular and extracurricular learning activities on students' choice of chiropractic technique. *J Chiropr Educ*. 2016;30(1):30–36.
14. Himelfarb I. A primer on standardized testing: history, measurement, classical test theory, item response theory and equating. *J Chiropr Educ*. 2019;33(2):151–163.
15. Himelfarb I, Shotts B, Tang N-E, Smith M. Score production and quantitative methods used by the NBCE for post-exam analyses. *J Chiropr Educ* 2020; 34(1):35–42.
16. Daniel WW. Biostatistics: *A Foundation for Analysis in the Health Sciences*. 7th ed. New York, NY: John Wily & Sons; 1999.
17. SPSS Inc. Armonk, NY: IBM SPSS Statistics for Windows; 2016.
18. Peugh JL, Enders CK. Missing data in educational research: a review of reporting practices and suggestions for improvement. *Rev Educ Res*. 2004;74:525–556.
19. Tabachnick BG, Fidell L. *Using Multivariate Statistics*. 7th ed. New York, NY: Pearson; 2018.
20. Pedhazur EJ. *Multiple Regression in Behavioral Research: Explanation and Prediction*. 3rd ed. Fort Worth, TX: Holt, Rinehart & Winston; 1997.

21. Bonferroni C. Teoria statistica delle classi e calcolo delle probabilità. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze. Libreria internazionale Seeber; 1936.
22. Crano WD, Brewer MB, Lac A. *Principles and Methods of Social Research*. 3rd ed. New York, NY: Routledge; 2014.
23. Kerlinger FN, Lee HB. *Foundations of Behavioral Research*. 4th ed. Orlando, FL: Harcourt Brace Jovanovich College Publishers; 2000.
24. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Erlbaum; 1988.
25. Kutner MH, Nechtsheim C, Neter J, Li W. *Applied Linear Statistical Models*. 5th ed. Boston, MA: McGraw-Hill; 2005.
26. Lewin LO, Singh M, Bateman BL, Glover PB. Improving education in primary care: development of an online curriculum using the blended learning model. *BMC Med Educ*. 2009;9(33):1–7.
27. Corbo M, Patel JP, Tawab A, Davies JG. Evaluating clinical skills of undergraduate pharmacy students using objective structured clinical examinations (OSCEs). *Pharm Educ*. 2006;6(1):1–6.
28. Lewin LO, Rapp KK, Hodder SL, et al. Performance of third-year primary-care-track students in an integrated curriculum at Case Western Reserve University. *Acad Med*. 74(1):82–89.
29. Patricio MF, Juliao M, Fareleira F, Vaz Carneiro A. Is the OSCE a feasible tool to assess competencies in undergraduate medical education? *Med Teach*. 2013; 35(6):503–514.